Chapter 2

Methodology

In the first report to Congress, *The Incidence and Severity of Sediment Contamination in Surface Waters of the United States: National Sediment Quality Survey* (USEPA, 1997), EPA noted that it faced two primary challenges in achieving the short-term goals of the *National Sediment Quality Survey* and fulfilling the mandate of the Water Resources Development Act (WRDA) of 1992, as described in the introduction to this report. Those two challenges still exist in this first update to the *National Sediment Quality Survey*. The first challenge was to compile a database of consistent sediment quality measures suitable for all regions of the country. The second was to identify scientifically sound methods to determine whether a particular sediment is "contaminated" based on the definition set forth in the statute.

In many known areas of contamination, visible and relatively easy-to-recognize evidence of harmful effects on resident biota is concurrent with elevated concentrations of contaminants in sediment. In most cases, however, less obvious effects on biological communities and ecosystems are much more difficult to identify and are frequently associated with varying concentrations of sediment contaminants. In other words, bulk sediment chemistry measures are not always indicative of toxic effect levels. Similar concentrations of a chemical can produce widely different biological effects in different sediments. This discrepancy occurs because toxicity is influenced by the extent to which chemical contaminants bind to other constituents in sediment. These other sediment constituents, such as organic ligands and inorganic oxides and sulfides, are said to control the bioavailability of accumulated contaminants (Di Toro et al., 1990). Toxicant binding, or sorption, to sediment particles suspends the toxic mode of action in biological systems (Swartz et al., 1995). Because the binding capacity of sediment varies, the degree of toxicity exhibited also varies for the same total quantity of toxicant.

The five general categories of sediment quality measurements are sediment chemistry, sediment toxicity, community structure, tissue chemistry, and pathology (Power and Chapman, 1992). Each of these categories has strengths and limitations for a national-scale sediment quality assessment. To be efficient in collecting usable data of similar types, EPA sought data that were available in electronic format, represented broad geographic coverage, and represented specific sampling locations identified by latitude and longitude coordinates.

As described previously, sediment chemistry measures alone might not accurately reflect risk to the environment. However, EPA has developed assessment methods that combine contaminant concentration with measures of the primary binding phase to address bioavailability for certain chemical classes, under assumed conditions of thermodynamic equilibrium (USEPA, 2000b). Other methods, which rely on statistical correlations of contaminant concentrations with incidence of adverse biological effects, also exist (Barrick et al., 1988; FDEP, 1994; Field et al., 1999; 2001 [in press] Long et al., 1995; MacDonald et al., 1996). In addition, fish tissue levels can be predicted using sediment contaminant concentrations, along with independent field measures of chemical partitioning behavior and other known or assigned fish tissue and sediment characteristics. EPA can evaluate risk to consumers from predicted and field-measured tissue chemistry data using established dose-response relationships and standard consumption patterns (USEPA and USACE, 1994). Evaluations based on tissue chemistry circumvent the bioavailability issue while also accounting for other mitigating factors such as metabolism. The primary difficulty in using field-measured tissue chemistry is relating chemical residue levels to a specific sediment, especially for those fish species which typically forage across great distances.

Sediment toxicity, community structure, and pathology measures are less widely available than sediment chemistry and fish tissue data in the broad-scale electronic format EPA sought for the NSI database. Traditionally sediment toxicity data have been expressed as the percentage survived in comparison to a control for indicator organisms exposed to the field-sampled sediment in laboratory

bioassays (USEPA, 1994a, b, 2000c). More recently, as indicated in the data collected for this report, sublethal measurements (e.g., reduction in survival, growth, and reproduction) are being used. These sublethal endpoints are more prevalent in this update to the first *National Sediment Quality Survey* report. Although these measures account for bioavailability and the antagonistic and synergistic effects of pollutant mixtures, they do not identify specific contaminants responsible for observed toxicity. Indicator organisms also might not represent the most sensitive species. Community structure measures, such as fish abundance and benthic diversity, and pathology measures are potentially indicative of long-term adverse effects, yet there are a multitude of mitigating physical, hydrologic, and biological factors that might not relate in any way to chemical contamination.

Studies have been conducted evaluating the ecological relevancy between response endpoints (i.e., reduction in growth of *Hyalella azteca*) and the ecological resources to be protected (i.e., the indigenous benthic community). Burton et al. (1996) compared results from laboratory sediment toxicity tests to colonization of artificial substrates exposed *in situ* to contaminated sediments. Survival and growth of *H. azteca* and *Chironomus tentans* in laboratory exposures negatively correlated to percent chironomids and percent tolerant taxa colonizing artificial substrates in the field. Schlekat et al. (1994) also reported general good agreement between sediment toxicity tests with *H. azteca* and benthic community responses.

An important goal of this report is to evaluate data collected throughout the United States in an attempt to describe the ecological integrity of sediments in the Nation's waterways. Ideally, the assessment methodology used to accomplish this task would be based on matched data sets of all five types of sediment quality measures described above to take advantage of the strengths of each measurement type and to minimize their collective weaknesses in a weight-of-evidence approach. Unfortunately, such a database does not exist on a national scale, nor is it typically available on a smaller scale. The statutory definition of contaminated sediments in WRDA 1992 enables EPA to identify locations where sediment chemistry measures exceed "appropriate geochemical, toxicological, or sediment quality criteria or measures." By the same statutory definition, based on screening values (e.g., EPA risk levels for fish tissue consumption) or availability of control samples for purposes of comparison, EPA can also use tissue chemistry and sediment toxicity measures to identify aquatic sediments that "otherwise pose a threat to human health or the environment." Without appropriate comparable reference conditions, EPA believes that it cannot accurately evaluate community structure or pathology measures to identify contaminated sediments based purely on the statutory definition.

For the first report to Congress the following measurement parameters and techniques were used alone or in combination to evaluate the probability of adverse effects:

Aquatic Life

(1) Comparison of sediment chemistry measurements to sediment chemistry screening values

- Draft sediment quality criteria (SQCs)
- Sediment quality advisory levels (SQALs)
- Effects range-median (ERM) and effects range-low (ERL) values
- Probable effects levels (PELs) and threshold effects levels (TELs)
- Apparent effects thresholds (AETs)

(2) Comparison of the molar concentration of acid volatile sulfides ([AVS]) in sediment to the molar concentration of simultaneously extracted metals ([SEM]) in sediment (under equilibrium conditions, sediment with [AVS] greater than [SEM] will not demonstrate toxicity from metals)

(3) Lethality based on sediment toxicity data

Human Health

(4) Comparison of theoretical bioaccumulation potential (TBP) values derived from sediment chemistry to:

- EPA cancer and noncancer risk levels or
- Food and Drug Administration (FDA) tolerance, action, or guidance values in the absence of, or if more stringent than, EPA levels

(5) Comparison of fish tissue contaminant levels to:

- EPA cancer and noncancer risk levels or
- FDA tolerance, action, or guidance values in the absence of, or if more stringent than, EPA levels

For the first report to Congress EPA evaluated more than 21,000 sampling stations nationwide as part of the NSI data evaluation. Of the sampling stations evaluated, 5,521 stations (26 percent) were classified as Tier 1, 10,401 (49 percent) were classified as Tier 2, and 5,174 (25 percent) were classified as Tier 3.

For the current analysis in this update, EPA evaluated sediment chemistry, tissue chemistry, and sediment toxicity data, taken at the same sampling station, individually and in combination using a variety of assessment methods. Because of the limitations of the available sediment quality measures and assessment methods, EPA characterizes this identification of contaminated sediment locations as a screening-level analysis. Similar to a potential human illness screen, a screening-level analysis should pick up potential problems and note them for further study. A screening-level analysis typically identifies many potential problems that prove not to be significant upon further analysis. Thus, classification of sampling stations in this analysis is not meant to be definitive, but is intended to be indicative of potential problems arising from persistent metal and organic chemical contaminants.

The first report to Congress used all data available from 1980 through 1993 for developing a baseline assessment. Because of the regular reporting requirements associated with this report, EPA wished to "window in" on a regular time frame for including data. The principal advantage of screening out older data (data collected prior to January 1, 1990) is to prevent the results presented in this report from being unduly influenced by historical data when more recent data are available. However, EPA recognizes that this "time windowing" will result in locations that have no evaluation provided in this document even though data are available in the NSI database. For the current analysis, EPA elected to evaluate data collected from 1990 to 1999 and to evaluate each chemical or biological measurement taken at a given sampling station individually. The methodology used for the current analysis has been modified to take advantage of scientific advances since the release of the first *National Sediment Quality Survey*. Similar to the previous analysis, sampling data obtained at a sampling station during the past 10 years for an individual chemical may result in the sampling station's being associated with adverse effect on either aquatic life or human health. The final section in Chapter 3 presents a comparison based on applying the methodology presented in this chapter to the data used for the first *National Sediment Quality Survey*.

EPA recognizes that sediment is dynamic and that great temporal and spatial variability in sediment quality exists. This variability can be a function of sampling (e.g., a contaminated area might be sampled one year, but not the next) or a function of natural events (e.g., floods can move contaminated sediment from one area to another, or can bury contaminated sediment). Movement of sediment is highly temporal and dependent upon the physical and biological processes at work in the watershed. Some deposits will redistribute while others will remain static unless disturbed by extreme events.

In this report, EPA associates sampling stations with their "probability of adverse effects on aquatic life or human health." Each sampling station falls into one of three categories (tiers): associated adverse effects on aquatic life or human health are probable (Tier 1); associated adverse effects on aquatic life and human health are possible (Tier 2); or no indication of associated adverse effects (Tier 3). A Tier 3 sampling station classification does not necessarily imply a zero or minimal probability of adverse effects,

only that available data (which may be substantial or limited) do not indicate an increased probability of adverse effects. Recognizing the imprecise nature of the numerical assessment parameters, Tier 1 sampling stations are distinguished from Tier 2 sampling stations based on the magnitude of a sediment chemistry measure or the degree of corroboration among the different types of sediment quality measures.

The remainder of this chapter presents a brief description of the NSI data, an explanation of the data evaluation approach, and the strengths and limitations of the data evaluation used for this *National Sediment Quality Survey*.

Description of NSI Data

The NSI database includes data from numerous data storage systems and monitoring programs. These systems and programs are listed below along with the percentage of stations that make up the NSI database.

- Selected data sets from EPA's Storage and Retrieval System (STORET) (35 percent of sampling stations)

  - U.S. Army Corps of Engineers (USACE)
  - EPA
  - States

- NOAA's Query Manager Data System (18.5 percent of sampling stations)

  - Including NOAA's National Status and Trends Program

- State of Washington Department of Ecology's Sediment Quality Information System (SEDQUAL) (16.5 percent of sampling stations)
- Selected data sets from USGS's WATSTORE (13.5 percent of sampling stations)
- EPA's Environmental Monitoring and Assessment Program (EMAP) (6.5 percent of sampling stations)
- Data compiled for the previous report to Congress (4.8 percent of sampling stations)
- Chesapeake Bay Program (2.4 percent of sampling stations)
- Upper Mississippi River System data compilation prepared by the USGS (1.1 percent of sampling stations)
- Other sampling programs (1.7 percent of sampling stations)

  - Indiana Department of Environmental Management Sediment Sampling Program
  - Oklahoma Reservoir Fish Tissue Monitoring Program, 1990-1998
  - Houston Ship Channel Toxicity Study

Although EPA elected to evaluate data collected since 1990 (i.e., 1990-1999), data before 1990 are maintained in the NSI database. At a minimum, EPA required that electronically available data include monitoring program, sampling date, latitude and longitude coordinates, and measured units for inclusion in the data evaluation. Additional information about available data fields is presented in Appendix A of this report.

The types of data contained in the NSI database include the following:

- Sediment chemistry: Measurement of the chemical composition of sediment-associated contaminants.
- Tissue residue: Measurement of chemical contaminants in the tissues of organisms.

- Toxicity: Measurement of the lethal or sublethal effects of contaminants in environmental media on various test organisms.

The NSI database represents a compilation of environmental monitoring data from a variety of sources. Most of the component databases are maintained under known and documented quality assurance and quality control procedures. However, EPA's STORET database is intended to be a broad-based repository of data. Consequently, the quality of the data in STORET, in terms of both database entry and analytical instrument error, is unknown and probably varies a great deal depending on the quality assurance management associated with specific data submissions.

Inherent in the diversity of data sources are contrasting monitoring objectives and scope. Component sources contain data derived from different spatial sampling plans, sampling methods, and analytical methods. For example, most data from EPA's EMAP program represent sampling stations that lie on a standardized grid over a given geographic area, whereas data in EPA's STORET most likely represent state monitoring data sampled from locations near known discharges or thought to have elevated contaminant levels. In contrast, many of NOAA's National Status and Trends Program data represent sampling stations purposely selected because they are removed from known discharges.

From an assessment point of view, STORET data might be useful for developing a list of contaminated sediment locations but might overstate the general extent of contaminated sediment in the Nation by focusing largely on areas most likely to be problematic. On the other hand, analysis of EMAP data might result in a more balanced assessment in terms of the mix of contaminated sampling stations and uncontaminated sampling stations. Approximately one-third of the sampling stations in the NSI database are from the STORET database. Reliance on these data is consistent with the stated objective of this survey: to identify those sediments which are contaminated. Realizing that uncontaminated areas are most likely substantially underrepresented, and the data in the NSI database does not provide a complete National coverage, EPA does not believe it is appropriate to make inferences regarding the overall condition of the Nation's sediment or characterizing the "percent contamination" using the data in the NSI database.

NSI data do not evenly represent all geographic regions in the United States as mentioned above, nor do the data represent a consistent set of monitored chemicals. For example, several of the databases are targeted toward marine environments or other geographically focused areas. Table 2-1 presents the number of stations evaluated per state. More than two-thirds of all stations evaluated in the NSI database are located in Washington, Virginia, California, Illinois, Florida, Wisconsin, New York, Texas, Oregon, and South Carolina. Each of these states has more than 500 monitoring stations. Other states of similar or larger size (e.g., Georgia, Pennsylvania) have far fewer sampling stations with data for evaluation. Figures 2-1, 2-2, and 2-3 depict the location of monitoring stations with data collected from 1990 through 1999 for sediment chemistry, tissue residue, and toxicity data, respectively. Individual stations may vary considerably in terms of the number of chemicals monitored. Some stations have data that represent a large number of organic and inorganic contaminants, whereas others have measured values for only a few chemicals. Thus, the inventory should not be construed as comprehensive even for locations with sampling data. The reliance on readily available electronic data has undoubtedly led to exclusions of a vast amount of information available from sources such as local and state governments and published reports. Other limitations, including data quality issues, are included in the Conclusions and Discussion chapter of this report.

Table 2-1. Number of Stations Evaluated in the NSI by State.

Region 1	Connecticut	121	Region 6	Arkansas	34
	Maine	0		Louisiana	396
	Massachusetts	127		New Mexico	167
	New Hampshire	4		Oklahoma	292
	Rhode Island	18		Texas	600
	Vermont	5			
Region 2	New Jersey	492	Region 7	Iowa	113
	New York	753		Kansas	119
	Puerto Rico	10		Missouri	194
	U.S. Virgin Islands	0		Nebraska	157
Region 3	Delaware	234	Region 8	Colorado	133
	District of Columbia	6		Montana	11
	Maryland	290		North Dakota	33
	Pennsylvania	216		South Dakota	32
	Virginia	1,577		Utah	56
	West Virginia	105		Wyoming	29
Region 4	Alabama	173	Region 9	Arizona	123
	Florida	1,157		California	1,535
	Georgia	263		Hawaii	18
	Kentucky	63		Nevada	76
	Mississippi	187			
	North Carolina	291			
	South Carolina	576			
	Tennessee	164			
Region 5	Illinois	1,370	Region 10	Alaska	290
	Indiana	233		Idaho	38
	Michigan	30		Oregon	599
	Minnesota	339		Washington	4,403
	Ohio	441			
	Wisconsin	777			

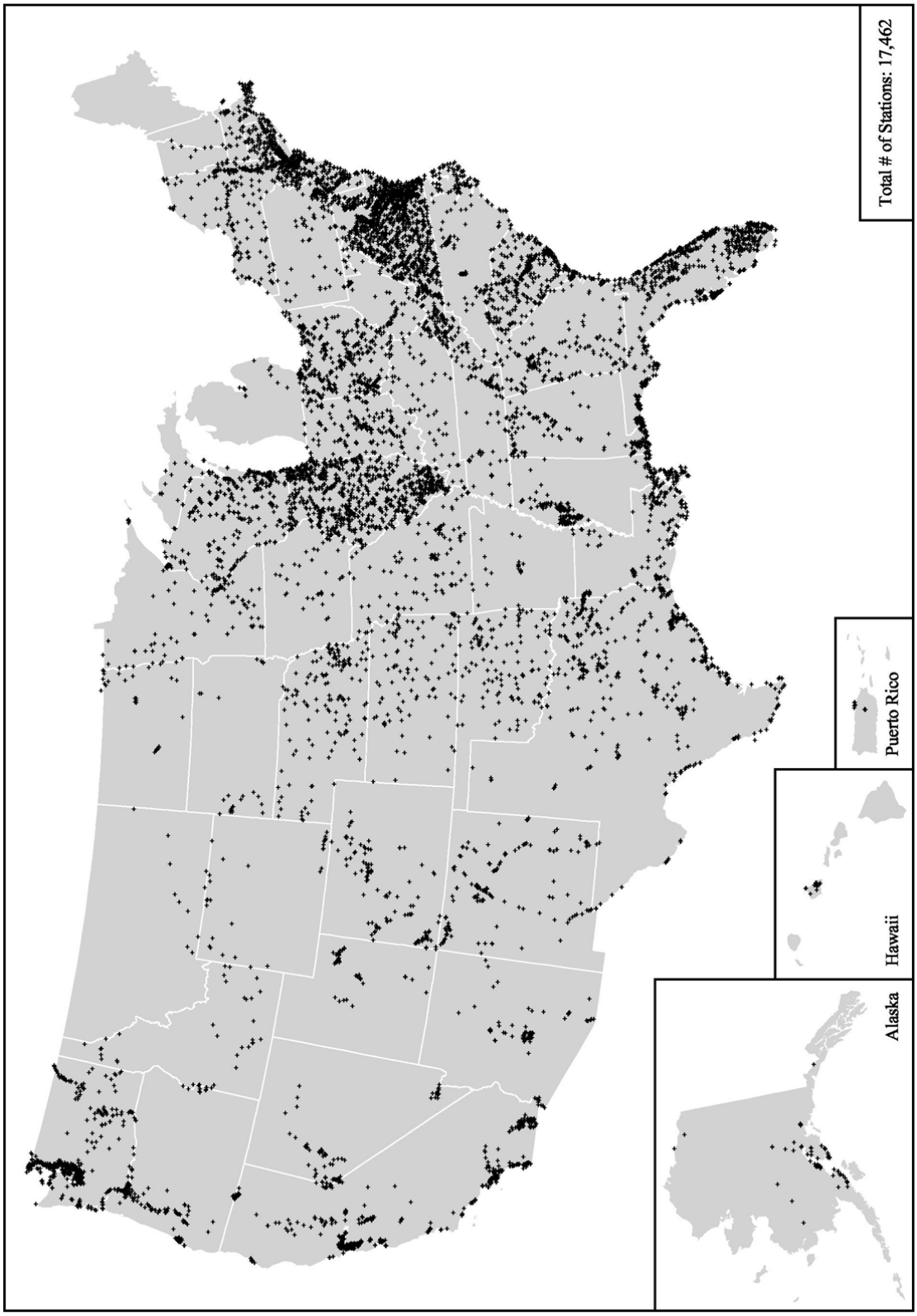

Figure 2-1. NSI Sediment Sampling Stations Evaluated.

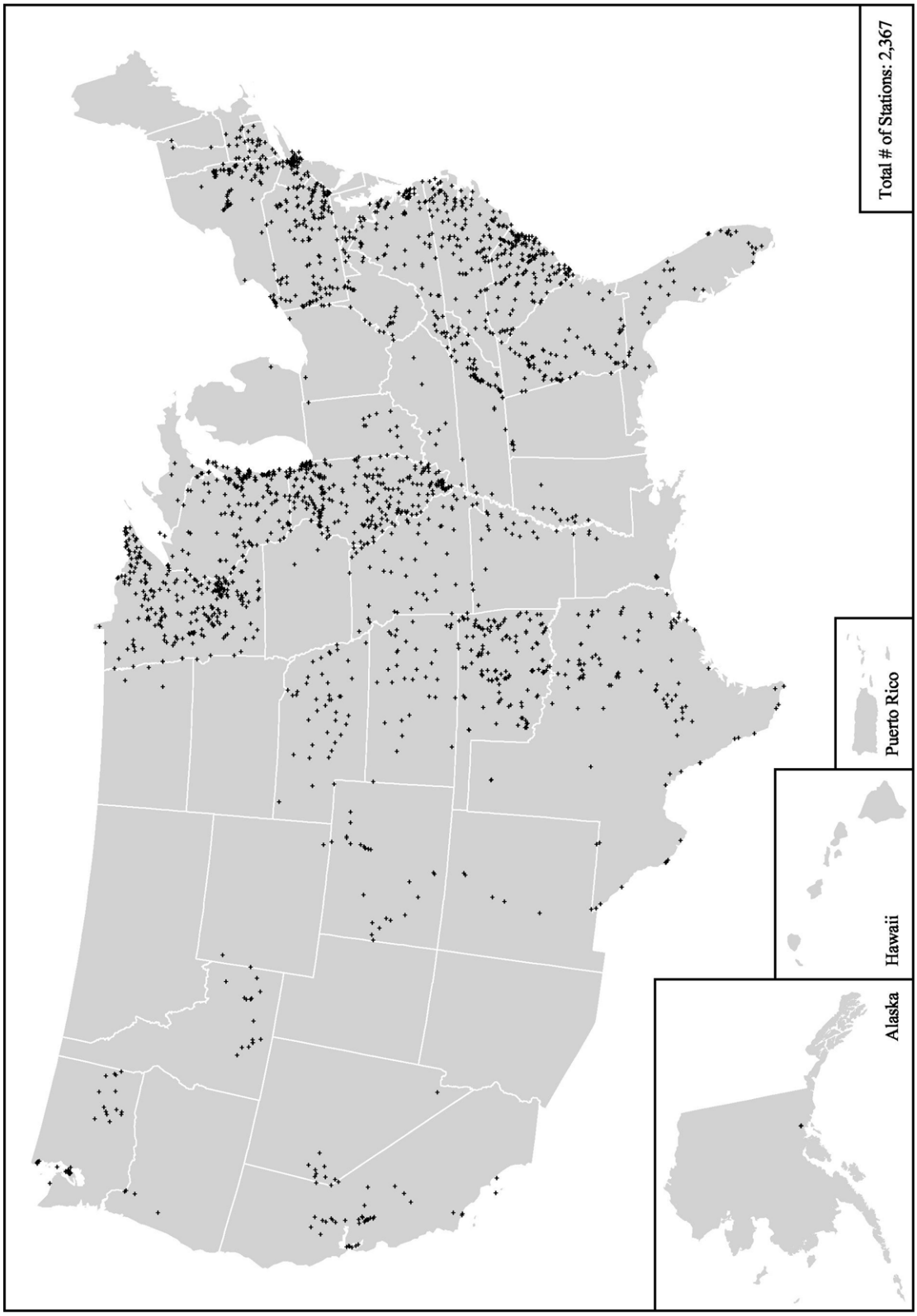

Figure 2-2. NSI Tissue Residue Sampling Stations Evaluated.

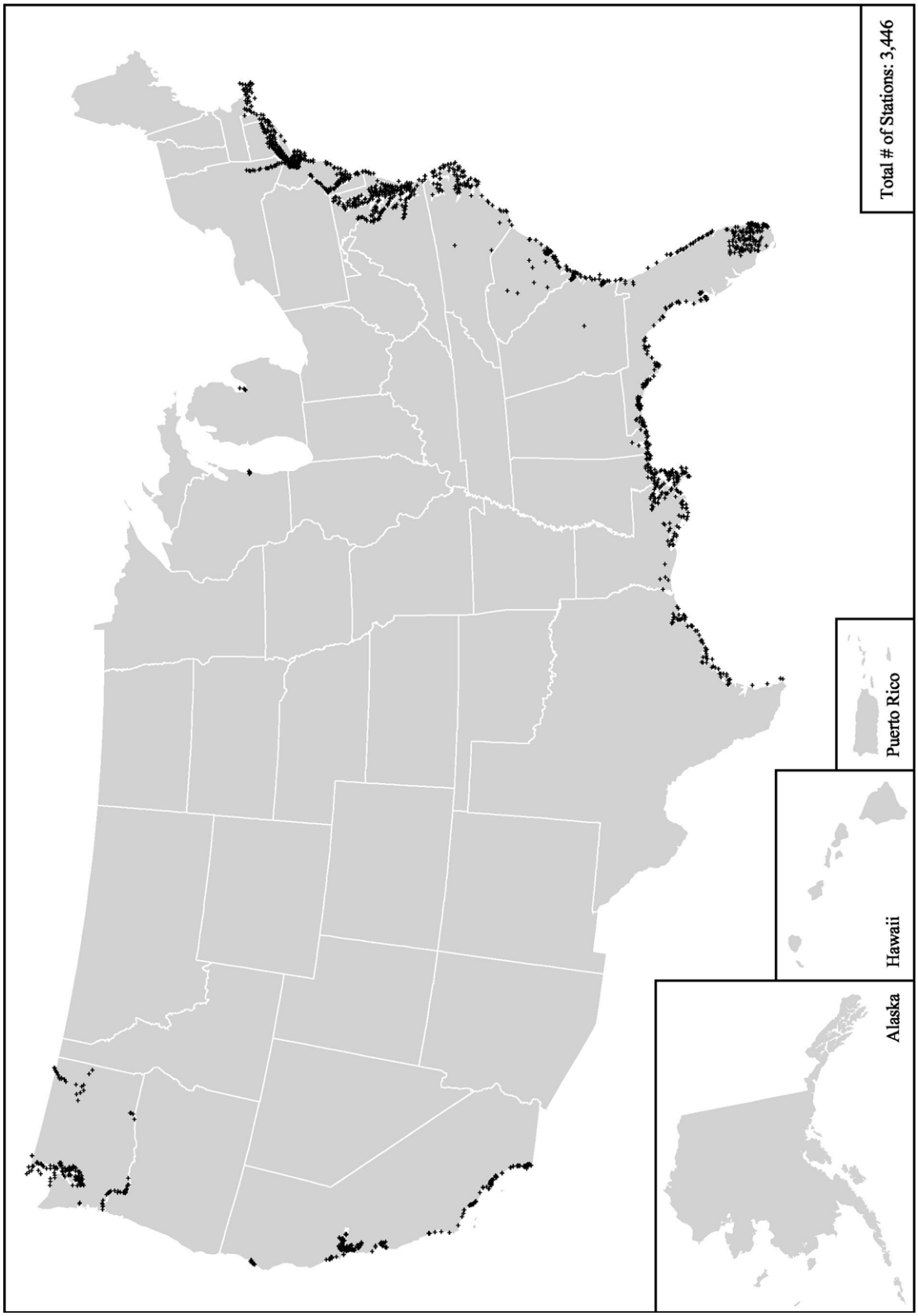

Figure 2-3. NSI Toxicity Test Stations Evaluated.

NSI Data Evaluation Approach

The methodology developed for this report for classifying sampling stations according to the probability of adverse effects on aquatic life and human health from sediment contamination relies on measurements of sediment chemistry, sediment toxicity, and contaminant residue in tissue. The approach used to evaluate the NSI data focuses on the protection of benthic organisms from exposure to contaminated sediments and the protection of humans from the consumption of fish that bioaccumulate contaminants from sediment. Table 2-2 presents the classification scheme used in the evaluation of the NSI data. Each component, or evaluation benchmark, of the classification scheme is numbered on Table 2-2. Each evaluation benchmark is discussed under a section heading cross-referenced to these numbers.

EPA analyzed the NSI data by evaluating each benchmark in Table 2-2 on a measurement-by-measurement and sampling station-by-sampling station basis. Each sampling station was associated with a "probability of adverse effects" by combining benchmarks as shown in Table 2-2. Because each individual measurement was considered independently except for divalent metals, polychlorinated biphenyls (PCBs), and DDT, whose concentrations were summed, and polycyclic aromatic hydrocarbons (PAHs), whose effect was analyzed as a mixture, a single observation of elevated concentration could place a sampling station into Tier 1 (associated with probable adverse effects). Any sampling station not meeting the requirements to be classified as Tier 1 or Tier 2 was classified as Tier 3. Sampling stations classified as Tier 3 include those for which substantial data were available without evidence of adverse effects, as well as sampling stations for which limited data were available to determine the potential for adverse effects.

Applying individual evaluation benchmarks to various measurements independently could lead to different site classifications. If one evaluation benchmark indicated Tier 1 but another evaluation benchmark indicated Tier 2 or Tier 3, a Tier 1 classification was assigned to the sampling station. For example, if a sampling station was categorized as Tier 2 based on all sediment chemistry data but was categorized as Tier 1 based on toxicity data, the station was placed in Tier 1. This principle also applies to evaluating multiple contaminants within the same evaluation benchmark. For example, if the evaluation of sediment chemistry data placed a sampling station in Tier 1 for PCBs and in Tier 2 for metals, the station was placed in Tier 1.

Recognizing the imprecise nature of some assessment benchmarks used in this report, Tier 1 sampling stations are distinguished from Tier 2 sampling stations based on the magnitude of a contaminant concentration in sediment or based on the degree of corroboration among the different types of sediment quality measures. This approach of integrating several assessment methods has been described as the most desirable approach for assessing the effects of contaminants associated with sediments (Ingersoll et al., 1996; 1997; Long and Morgan, 1990; MacDonald et al., 1996; USEPA 2000c). In response to uncertainty in both biological and chemical measures of sediment contamination, environmental managers must balance Type I errors (false positives: sediment classified as posing a threat when in fact it does not) with Type II errors (false negatives: sediment that poses a threat but was not classified as such). In screening analyses, the environmentally protective approach is to minimize Type II errors, which leave toxic sediment unidentified. To achieve a balance and to direct attention to areas most likely to be associated with adverse effects, Tier 1 sampling stations are intended to have a high rate of "correct" classification (e.g., sediment definitely posing or definitely not posing a threat) and a balance between Type I and Type II errors. On the other hand, to retain a sufficient degree of environmental conservatism in screening, Tier 2 sampling stations are intended to have a very low number of false negatives in exchange for a large number of false positives.

For this NSI data evaluation EPA opted to analyze data collected since 1990 with valid latitude and longitude coordinates. The numbered evaluation benchmarks used in the NSI data evaluation are briefly described below. A detailed description of the evaluation benchmarks is presented in Appendix B.

Table 2-2. NSI Data Evaluation Approach

Sampling Station Classification	Data Used to Determine Classifications				
	Sediment Chemistry		**Tissue Residue**		**Toxicity**
Tier 1: Associated Adverse Effects on Aquatic Life or Human Health Are Probable	Sediment chemistry value exceeds a draft equilibrium partitioning sediment guideline (ESG) derived from a final or secondary acute value (FAV or SAV)[a] **1**	**OR**	Tissue levels of chemicals with a log $K_{OW} \geq 5.5$ in samples[d] that exceed EPA's human health cancer risk of 10^{-5}, a noncancer HQ of 1, or FDA's Guidance/Action/Tolerance levels **12**	**OR**	Toxicity demonstrated by one solid-phase sediment test resulting in (1) < 75% control-adjusted survival, (2) freshwater invertebrate (*Hyalella azteca*) sublethal toxicity < 90% control-adjusted length, or (3) freshwater invertebrate (*Hyalella azteca*, *Chironomus tentans,* and *Chironomus riparius*) sublethal toxicity < 70% control-adjusted weight **15**
	OR [SEM]-[AVS] > 5 for the sum of molar concentrations of Cd, Cu, Ni, Pb, Zn, and ½ x Ag[b] **2**				
	OR Any sample with a predicted proportion toxic ≥ 0.5 using a logistic regression model **3**				**OR** Any sample meeting the benchmark described under Tier 2 for toxicity using at least two different species **16**
	OR Sum PAH ESG toxicity unit (draft) derived from FAV > 1[ac] **4**				
	OR Sediment chemistry TBP exceeds EPA's human health cancer risk of 10^{-4} or a noncancer hazard quotient (HQ) of 1[a] **5**				
	OR For chemicals with log $K_{OW} < 5.5$, sediment chemistry TBP exceeds (EPA's) human health cancer risk of 10^{-5}, a noncancer HQ of 1, or FDA's Guidance/Action/Tolerance levels[a] **6**	**AND**	Tissue levels of chemicals with a log $K_{OW} < 5.5$ in samples[d] that exceed EPA's human health cancer risk of 10^{-5}, a noncancer HQ of 1, or FDA's Guidance/ Action/Tolerance levels **13**	—	—
Tier 2: Associated Adverse Effects on Aquatic Life or Human Health Are Possible	Sediment chemistry value exceeds a draft ESG derived from a final or secondary chronic value (FCV or SCV)[a] **7**	**OR**	Tissue levels of chemicals with a log $K_{OW} < 5.5$ in samples[d] that exceed EPA's human health cancer risk of 10^{-5}, a noncancer HQ of 1, or FDA's Guidance/ Action/Tolerance levels **14**	OR	Toxicity demonstrated by one solid-phase sediment test resulting in (1) < 90% control-adjusted survival (but ≥ 75% control-adjusted survival), (2) freshwater invertebrate (*Hyalella azteca*) sublethal toxicity < 95% control-adjusted length (but ≥ 90% control-adjusted length), or (3) freshwater invertebrate (*Hyalella azteca*, *Chironomus tentans, and Chironomus riparius*) sublethal toxicity < 90% control-adjusted weight (but ≥70% control-adjusted weight) **17**
	OR [SEM]-[AVS] = 0-5 for the sum of molar concentrations of Cd, Cu, Ni, Pb, Zn, and ½ x Ag[b] **8**				
	OR Any sample with a predicted proportion toxic ≥ 0.25 but < 0.5 using a logistic regression model **9**				
	OR Sum PAH ESG toxicity unit (draft) derived from FCV > 1[ac] **10**				
	OR Sediment chemistry TBP exceeds EPA's human health cancer risk of 10^{-5}, a noncancer HQ of 1, or FDA's Guidance/Action/Tolerance levels **11**				
Tier 3: No Indication of Associated Adverse Effects	Any sampling station not categorized as Tier 1 or Tier 2. Available data (which may be very limited or quite extensive) do not indicate a likelihood of adverse effects on aquatic life or human health.				

[a] If total organic carbon (TOC) is not reported, a default value of 1% was assumed. For ESG-based methods if the reported TOC is less than 0.2%, a default TOC value of 0.2% was used.

[b] Metals: Cd = cadmium, Cu = copper, Ni = nickel, Pb = lead, Zn = zinc, Ag = silver.

[c] Acenaphthene, acenaphthylene, anthracene, benzo(a)anthracene, benzo(a)pyrene, benzo(b)fluoranthene, benzo(k)fluoranthene, chrysene, fluoranthene, fluorene, naphthalene, phenanthrene, and pyrene used to compute ESG toxicity unit.

[d] Only those species considered benthic (demersal), nonmigratory (resident), and edible by human populations are included in human health assessments.

As is noted in the first footnote to Table 2-2, if the total organic carbon (TOC) was not reported, a default of 1 percent was assumed. This assumption was based on a literature review performed during the preparation of the first report to Congress. TOC values can range from 0.1 percent in sandy sediments to 1 to 4 percent in silty harbor sediments and 10 to 20 percent in navigation channel sediments (Clarke and McFarland, 1991). Long et al. (1995) reported an overall mean TOC concentration of 1.2 percent from data compiled from 350 publications for their biological effects database for marine and estuarine sediments. Ingersoll et al. (1996) reported a mean TOC concentration of 2.7 percent for inland freshwater samples. Based on this review of TOC data, EPA selected a default TOC value of 1 percent for this evaluation. Consistent with the screening level application, this value should not lead to an underestimate of the bioavailability of associated contaminants in most cases.

Sediment Chemistry Data

The sediment chemistry screening values used in this report are not regulatory criteria, site-specific clean-up standards, or remediation goals. Sediment chemistry screening values are reference values above which a sediment ecotoxicological assessment might indicate a potential threat to aquatic life. The sediment chemistry screening values used to evaluate the NSI data for potential adverse effects of sediment contamination on aquatic life include values based on theoretical calculations and empirically/statistically derived values. The theoretically based values rely on the physical/chemical properties of sediment and chemicals to predict the level of contamination that would not cause an adverse effect on aquatic life. The empirically/statistically derived screening values are based on estimating the probability that a sediment toxicity test would indicate significant toxicity using multiple chemical measures of 37 target chemicals.

The theoretically based screening values used in the evaluation of NSI data include five different draft Equilibrium Partitioning Sediment Guidelines (ESGs) developed by EPA. Draft ESGs were developed for dieldrin, endrin, 32 nonionic organics, mixtures of PAHs, and metal mixtures. The use of each of these screening values in the evaluation of the NSI data is described below. Another theoretically based evaluation benchmark, the theoretical bioaccumulation potential (which was used for human health assessments), is also described below.

Sediment Chemistry Values Exceed EPA Draft Equilibrium Partitioning Sediment Guideline (ESG) [1,7]

EPA developed draft ESGs using the equilibrium partitioning (EqP) approach (described in detail in Appendix B) for linking bioavailability to toxicity. This approach accounts for the varying biological availability of chemicals in different sediments and permits the incorporation of the relevant biological effects concentration. This approach enables the derivation of a guideline that is causally linked to the specific chemical, is applicable across sediments, and is protective of benthic organisms. The EqP theory asserts that a nonionic chemical in sediment partitions between sediment organic carbon, interstitial (pore) water, and benthic organisms. At equilibrium, if the concentration in any one phase is known, the concentration in the others can be predicted. EPA has developed different draft ESGs based on final or secondary acute or chronic values to reflect the differing degrees of data availability and uncertainty. These draft ESGs are expressed as a concentration of a chemical in sediment and are derived to protect aquatic benthic organisms from direct toxicity due to that chemical (or chemicals in the case of metals mixtures and PAH mixtures). The draft ESG for nonionic organics applies only to sediments that have at least 0.2 percent organic carbon. For samples with total organic carbon (TOC) less than 0.2 percent, a default TOC value of 0.2 percent was used.

Comparison of AVS to SEM Molar Concentrations [2, 8]

The use of the total concentration of a trace metal in sediment as a measure of its toxicity and its ability to bioaccumulate is problematic because different sediments exhibit different degrees of bioavailability for the same total quantity of metal (Di Toro et al., 1990; Luoma, 1983). These

differences have been reconciled by relating organism toxic response (mortality) to the metal concentration in the sediment interstitial water (Adams et al., 1985; Di Toro et al., 1990). Acid-volatile sulfide (AVS) is one of the major chemical components that control the activities and availability of metals in interstitial waters of anoxic (lacking oxygen) sediments (Meyer et al., 1994).

A large reservoir of sulfide exists as iron sulfide in anoxic sediment. Sulfide reacts with several divalent transition metal cations (cadmium, copper, mercury, nickel, lead, zinc) and predominantly monovalent silver to form highly insoluble compounds that are not bioavailable (Allen et al., 1993, Ankley et al., 1991, Berry et al., 1999, Carlson et al. 1991). It follows in theory, and with verification (Di Toro et al., 1990), that divalent transition metals will not begin to cause toxicity in anoxic sediment until the reservoir of sulfide is used up (i.e., the molar concentration of metals exceeds the molar concentration of sulfide), typically at relatively high dry-weight metal concentrations. This observation has led to a laboratory measurement technique of calculating the difference between simultaneously extracted metal (SEM) concentration and acid-volatile sulfide concentration from field samples to determine potential toxicity (Ankley et al., 1991, Carlson et al. 1991).

To evaluate the potential effects of metals on benthic species, the molar concentration of AVS ([AVS]) was compared to the sum of SEM molar concentrations ([SEM]) for six metals: cadmium, copper, nickel, lead, zinc, and silver. Molar concentrations of cadmium, copper, nickel, lead, and zinc are comparable with AVS on a one-to-one basis. Since silver exists predominantly as a monovalent metal, half the molar concentration of silver is compared with the molar AVS concentration. Mercury was excluded from AVS comparison because other important factors play a major role in determining the bioaccumulation potential of mercury in sediment. Specifically, under certain conditions mercury binds to an organic methyl group and is readily taken up by living organisms.

Sediment with measured [SEM] in excess of [AVS] does not necessarily exhibit toxicity. This is because other binding phases can tie up metals. However, research indicates that sediment with [AVS] in excess of [SEM] will not be toxic from metals, and the greater the [SEM]-[AVS] difference, the greater the likelihood of toxicity from metals. Analysis of toxicity data for freshwater and saltwater sediment amphipods (crustaceans) from EPA's Environmental Research Laboratory in Narragansett, Rhode Island, revealed that 80 to 90 percent of the sediments were toxic at [SEM]-[AVS] > 5 (Hansen, 1995; see also Hansen et al., 1996a, b). Thus, EPA selected [SEM]-[AVS] = 5 as the demarcation line between Tier 1 and Tier 2. For the purpose of this evaluation, where [SEM]-[AVS] was greater than 5, the sampling station was classified as Tier 1. If [SEM]-[AVS] was between zero and 5, the sampling station was classified as Tier 2. If [SEM]-[AVS] was less than zero, or if AVS or the six AVS metals were not measured at the sampling station, the sampling station was classified as Tier 3 unless otherwise classified by another benchmark.

There are several important factors to consider in interpreting the [SEM]-[AVS] difference. First, all toxic SEMs present in amounts that contribute significantly to the [SEM] sum should be measured. However, because mercury presents special problems, it is not included in the current SEM analysis. Second, if the AVS content of sediment is low, as in fully oxidized sediments, the metal-binding capacity of the sediment decreases and the method will not work (Adams et al., 1992; Zhuang et al., 1994). Most benthic macroorganisms, including those used in toxicity tests, survive in sediments that have a thin oxidized surface layer and then an anoxic layer. The anoxic layer can have significant AVS concentrations that would reduce the metal activity to which these organisms are exposed (Di Toro et al., 1992). Third, AVS varies spatially in sediment—vertically with depth and horizontally where patches of an appropriate carbon source occur under low-oxygen conditions for the sulfate-reducing bacteria. Finally, AVS can vary when sediments are oxygenated during physical disturbance and seasonally as changes in the productivity of the aquatic ecosystem alter the oxidation state of sediment and oxidize metal sulfides; therefore, the toxicity of the metals present in the sediment also changes over time (Howard and Evans, 1993).

Selection of an [SEM]-[AVS] difference sufficiently high to place a sediment in the Tier 1 classification requires careful consideration because the relationship between organism response and the [SEM]-[AVS] difference of sediment depends on the amount and kinds of other binding phases present. Using freshwater and saltwater sediment amphipod toxicity data, researchers at EPA's Environmental Research Laboratory in Narragansett, Rhode Island, plotted [SEM]-[AVS] versus the percentage of sediments with a higher [SEM]-[AVS] value that were toxic. For this analysis, the researchers defined toxicity as greater than 24 percent mortality. Analysis of these data reveals that between 80 percent and 90 percent of the sediments were toxic at [SEM]-[AVS] = 5. The running average mortality at this level was between 44 percent and 62 percent (Hansen, 1995). EPA's Office of Science and Technology selected [SEM]-[AVS] = 5 as the demarcation line between the higher (Tier 1) and intermediate (Tier 2) probability categories.

Predicted Proportion Toxic from Sediment Chemistry [3, 9]

The empirically based or correlative screening values used in the previous NSI data evaluation rely on paired field and laboratory data to relate incidence of observed biological effects to the dry-weight sediment contamination of a specific chemical. The empirically based, correlative screening values include the effects range-median (ERM)/effects range-low (ERL) values, probable effects level (PEL)/threshold effects level (TEL), and apparent effects thresholds (AET). Field et al. (1999, 2001 [in press]) developed an alternative method for the evaluation of sediment quality by using a logistic regression model. This model (described in detail in Appendix B) is used to predict the probability of observing specific toxic effects—for selected toxicity test endpoints and a wide range of concentrations—for individual contaminants. Using the sediment chemistry and toxicity data, individual logistic models were developed for each contaminant, and the slope and intercept values were calculated using the maximum likelihood approach.

A total of 37 chemicals are included in the logistic regression model. For the NSI data evaluation, the probability of toxic effects was computed for the various contaminants from individual logistic regression equations. The predicted proportion toxic was then estimated from the maximum probability of toxic effects using a regression equation. When the maximum predicted proportion toxic for any sample was $\geq$ 0.5, the sampling station was assigned to Tier 1. When the maximum predicted proportion toxic was $\geq$ 0.25 but < 0.5, the sampling station was classified as Tier 2. Other sampling stations with available data for chemicals included in the logistic regression model were classified as Tier 3 unless otherwise classified by another benchmark.

PAH-Based ESG Toxicity Unit Exceed Screening Benchmark [4, 10]

The $\Sigma ESGTU_{PAH}$ model estimates the probability of toxic effects in PAH-contaminated sediments by using equilibrium partitioning, the quantitative structure activity relationship (QSAR) technique, toxic unit, additivity, and concentration-response models (Swartz et al.,1995, Swartz, 1999). The model predicts the probability of acute sediment toxicity to marine and estuarine amphipods caused by a combination of PAHs. EPA's draft equilibrium partitioning sediment guideline (ESG) recommends an approach for summing the toxicological contributions of mixtures of 34 PAHs in sediments to determine whether their concentrations in any specific sediment are acceptable for the protection of benthic organisms from PAH toxicity. Because PAHs occur in sediments as mixtures and their toxicities in water, tissues, and sediments are additive or nearly additive (Di Toro, 2000), the consideration of their toxicities on an individual basis would result in guidelines that are under-protective. For this reason, EPA recommends the use of combined toxicological contributions of the PAH mixture in evaluating sediments.

Because many monitoring and assessment efforts measure a smaller group of PAHs, such as 13 or 23 PAHs, EPA has recommended adjustment factors to relate these smaller subsets to the expected concentration of the 34 PAHs. The total Equilibrium Partitioning Sediment Guideline Toxic Unit (ΣESGTU)—based on either final chronic or acute value—is used to classify sampling stations as either Tier 1 or Tier 2 (described in detail in Appendix B). For use in determining the uncertainty in predicting

$\Sigma ESGTU_{FCV,TOT}$ from datasets consisting of 13 or 23 PAHs, EPA combined two data sources that measured the 34 PAHs and treated the dataset as a single data source. In doing this data combination, a dataset containing both alkylated and parent PAHs with their correlative relationships was generated. Based on the relative distributions of the $\Sigma ESGTU_{FCV,TOT}$ to the $\Sigma ESGTU_{FCV}$ for the 13 PAHs, EPA recommended various multiplication factors to achieve various degrees of confidence levels. Table B-3 in Appendix B presents the relative distribution of the multiplication factors. The NSI data evaluation targeted 13 PAHs and used the EPA-recommended multiplication factor of 2.75 to obtain an accurate estimation of the ΣESGTU. However, for this data evaluation not all 13 PAHs were required to be measured at any one station for that station to be considered for tier classification. Based on the sensitivity analysis done, it was observed that this variation from the EPA recommended practice did not dramatically change the total number of station tier classification. This analysis applies only to sediments that have at least 0.2 percent organic carbon. For samples with total organic carbon (TOC) less than 0.2 percent, a default TOC value of 0.2 percent was used.

Sediment Chemistry TBPs Exceed Screening Benchmark [5, 6, 11]

This evaluation benchmark addresses the risk to human consumers of organisms exposed to sediment contaminants. The theoretical bioaccumulation potential (TBP) is an estimate of the equilibrium concentration (concentration that does not change with time) of a contaminant in tissues if the sediment in question were the only source of contamination to the organism. At present, the TBP calculation can be performed only for nonpolar organic chemicals. The TBP is estimated from the concentration of contaminant in the sediment, the organic carbon content of the sediment, the lipid content of the organism, and the relative affinity of the chemical for sediment organic carbon and animal lipid content. This relative affinity is measured in the field and is called a biota-sediment accumulation factor (BSAF). In practice, field-measured BSAFs can vary by an order of magnitude or greater for individual compounds depending on location and time of measurement. For this evaluation, EPA selected BSAFs that represent the central tendency, suggesting an approximate 50 percent chance that an associated tissue residue level would exceed a screening risk value.

In the evaluation of NSI data, if a calculated sediment chemistry TBP value exceeded a screening value derived using the EPA risk assessment methodology (i.e., EPA's human health cancer risk of 10^{-4} or a noncancer hazard quotient [HQ] of 10, evaluation benchmark 5), the station was classified as Tier 1. Individual chemical risk levels were considered separately; that is, risks from multiple contaminants were not added.

For chemicals with an octanol-water partition coefficient ($\log K_{ow}$) < 5.5, the following benchmark was used: If a calculated sediment chemistry TBP value exceeded a screening value derived using EPA's human health cancer risk of 10^{-5} or a noncancer hazard quotient [HQ] of 1 or the Food and Drug Administration's (FDA) tolerance/action/guidance level (evaluation benchmark 6), and if a corresponding tissue residue level for the same chemical in demersal, resident, and edible species at the same sampling station also exceeded one of those screening values (evaluation benchmark 13), the station was classified as Tier 1. Individual chemical risk levels were considered separately; that is, risks from multiple contaminants were not added. In this assessment, both sediment chemistry and tissue residue samples must have been taken from the same sampling station. If tissue residue levels for the same chemical for a demersal, resident and edible species at the same sampling station did not exceed standard EPA risk levels or FDA levels or there were no corresponding tissue data, the sampling station was classified as Tier 2.

In addition, for all chemicals irrespective of their octanol-water partition coefficient, when sediment chemistry TBP exceeded stated EPA risks or FDA guidelines shown in Table 2-2, the sample stations were classified as Tier 2. If neither TBP values nor fish tissue residue levels exceeded the appropriate EPA risk levels given in Table 2-2 or the FDA guidance levels, or if no chemicals with TBP values, EPA risk levels, or FDA levels were measured, the sampling station was classified as Tier 3 unless otherwise classified by another benchmark. A detailed description of the methods used to develop TBP values and to determine the EPA risk levels used in this comparison is presented in Appendix B.

Tissue Residue Data [12, 13, 14]

Tissue residue data were used to assess potential adverse effects on humans from the consumption of fish that become contaminated through exposure to contaminated sediment. Only those species considered benthic, nonmigratory (resident), and edible by human populations were included in human health assessments. A list of species included in the NSI database and their characteristics is presented in Appendix D.

For chemicals with a log $K_{ow} \geq 5.5$, if the tissue residue levels in demersal, resident, and edible species exceeded EPA risk screening values (i.e., EPA's human health cancer risk of 10^{-5} or a noncancer hazard quotient [HQ] of 1 or the FDA tolerance/action/guidance level), the station was classified as Tier 1.

For chemicals with a log $K_{ow} < 5.5$, both a tissue residue level exceeding an FDA tolerance/action/guidance level or stated EPA risk level and a sediment chemistry TBP value exceeding that risk/tolerance level for the same chemical were required to classify a sampling station as Tier 1. If tissue residue levels exceeded FDA levels or EPA risk levels but corresponding TBP values were not exceeded at the same station (or there were no sediment chemistry data from that station), the sampling station was classified as Tier 2. If neither fish tissue levels nor TBP values exceeded EPA risk levels or FDA levels, or if no chemicals with TBP values, EPA risk levels, or FDA levels were measured, the sampling station was classified as Tier 3 unless otherwise classified by another benchmark.

Toxicity Data [15, 16, 17]

Toxicity data were used to classify sediment sampling stations based on short- or long-term sediment toxicity tests. Nonmicrobial sediment toxicity tests based on survival and on variation in length or weight were evaluated. For all of the endpoints (i.e., survival and variations in length or weight), the test results were "adjusted" to compare against a control test for the same species (described in more detail in Appendix B). Toxicity test results that lacked control data were excluded. EPA has standardized testing protocols for marine and freshwater toxicity tests.

For the NSI data evaluation, only solid-phase bulk sediment toxicity tests, with test durations of seven or more days, were considered. Calculated values of the percentage of species surviving were reported by individual databases. These percentages were based on values adjusted for a control sample. Sampling stations with tests resulting in less than 75 percent of the control-adjusted survival in marine and freshwater species were classified as Tier 1. Similar to the results reported for percentage survival, calculated values of the percentage variation in length and weight were reported in various studies. These percentage values were also reported as adjusted for a control test. Sample stations with freshwater invertebrates (*Hyalella azteca*) that indicated sublethal toxicity by lengths of less than 90 percent of the control-adjusted length or with freshwater invertebrates (*Hyalella azteca*, *Chironomous tentans*, and *Chironomus riparius*) that indicated sublethal toxicity by weights of less than 70 percent of the control-adjusted weight were classified as Tier 1.

Stations were classified as Tier 2 based on benchmarks similar to those established for Tier 1 classification, but with lower threshold values. Toxicity tests resulting in less than 90 (but ≥ 75) percent of the control-adjusted survival for both marine and freshwater species were classified as Tier 2. Sampling stations with freshwater invertebrates (*Hyalella azteca*) that indicated sublethal toxicity by lengths of less than 95 (but ≥ 90) percent of the control-adjusted length or with freshwater invertebrates (*Hyalella azteca*, *Chironomus tentans*, and *Chironomus riparius*) that indicated sublethal toxicity by weights of less than 90 (but ≥ 70) percent of the control-adjusted weight were classified as Tier 2.

A station could be classified as Tier 2 by the benchmark stated above based on more than one test species. When a station was classified as Tier 2 based on results from two or more species from that station, the tier classification for that station was upgraded to Tier 1.

Strengths of the NSI Data Evaluation

For this report to Congress, EPA has compiled the most extensive database of sediment quality information currently available in electronic format. To evaluate these data, EPA has applied sediment assessment techniques in a weight-of-evidence approach recommended by national experts. The evaluation approach uses sediment chemistry, tissue residue, and toxicity test results. The assessment tools employed in this analysis have been applied in North America with results published in peer-reviewed literature. Toxicity test data were generated using established standard methods employed by multiple federal agencies. The evaluation approach addresses potential impacts on both aquatic life and human health.

Because of the complex nature of the reactions among different chemicals in different sediment types, in water, and in tissues, no single sediment assessment technique can be used to adequately evaluate potential adverse effects from exposure to all contaminants. Uncertainties and limitations are associated with all sediment quality evaluation techniques. To compensate for those limitations, EPA has used multiple assessment techniques, singularly and in combination, to evaluate the NSI data. For example, EPA applied draft equilibrium partitioning sediment guidelines for nonionic organics, for mixtures of polycyclic aromatic hydrocarbons, and for five divalent metals. The screening values used to evaluate the NSI data include both theoretical and correlative approaches. The theoretical approaches (e.g., draft ESGs and TBPs) are based on the best information available concerning how chemicals react in sediments and organisms and how organisms react to those chemicals. The correlative approach (i.e., logistic model) is based on matched sediment and biological data gathered in the field and in the laboratory, and it provides substantial evidence of actual biological effects from sediments.

The NSI data evaluation approach includes assessments of potential impacts on both human health and aquatic life. Some chemicals pose a greater risk to human health than to aquatic life; for others, the reverse is true. By evaluating both potential human health and aquatic life impacts, EPA has ensured that the most sensitive endpoint is used to assess environmental impacts.

Because sediment chemistry data are not the only indicators of potential environmental degradation due to sediment contamination, the NSI data evaluation approach also includes evaluations of fish tissue residue and toxicity data. If high levels of PCBs, dioxins, or other highly hydrophobic organic chemicals (commonly found associated with sediments) were measured in fish tissue at a given sampling station, the station could be categorized as Tier 1 with no corroborating sediment chemistry data. For other chemicals, high concentrations in tissues alone were not sufficient to categorize a sampling station as Tier 1; corroborating sediment chemistry data were also required. For a sampling stations to be categorized as Tier 1 based on toxicity data alone, only solid-phase tests were analyzed.

Limitations of the NSI Data Evaluation

This methodology was designed for the purpose of a screening-level assessment of sediment quality. A considerable amount of uncertainty is associated with the site-specific measures, assessment techniques, exposure scenarios, and default parameter selections. Therefore, the results of evaluating particular sampling stations based on this methodology should be followed up with more intensive assessment efforts, when appropriate (e.g., for waterbodies with multiple Tier1 sampling stations located in APCs). Two types of limitations are associated with the evaluation of the NSI data: limitations associated with the data themselves and limitations associated with the evaluation of the data.

Limitations of Data

The NSI database is a multimedia compilation of environmental monitoring data obtained from a variety of sources, including state and federal government offices. Inherent in the diversity of data sources are contrasting monitoring objectives and scopes, which make comparison of data from different data sets difficult. For example, several of the databases contain only information from marine

environments or other geographically focused areas. The potential for inconsistencies in measured concentrations of contaminants at different stations exists for samples taken from different monitoring programs. For example, sampling different age profiles in sediments, applying different sampling and analysis methods, and sampling for different objectives can affect the results of the NSI data evaluation. The surficial samples analyzed in this report vary since many different sampling devices are used depending water depth and study objectives. For this report, samples were included when the reported lower depth was no greater than 30 cm and the reported upper depth was null or less than 2 cm. Although some monitoring programs identified sampling and laboratory methods, these data are rarely provided with the data. In addition, some data sets included in the NSI database were not peer-reviewed (e.g., some data sets from EPA's STORET). Furthermore, each monitoring program used unique sampling and analysis protocols. For example, PCBs are measured by nearly all of the monitoring programs but were analyzed and reported as aroclor-specific data, congener-specific data, total PCBs, or a combination of these.

The only quality assurance/quality control (QA/QC) information required for data to be included in the NSI database was information on the source of the data and the location of the sampling station. Available information on several types of QA/QC procedures that can influence the quality of the data and can be used to check the quality of data was included in the NSI database. None of this information was required before a data set could be included in the NSI database; however, most of the component databases are maintained under known and documented QA/QC procedures. For the 19,470 stations evaluated in this report, approximately 97 percent contain sufficient information in the database to allow the user to contact an agency, contact an investigator, or reference a report to obtain the available QA information. Data reporting was also inconsistent among the different data sources. Inconsistencies that required resolution included the lack or inconsistent use of Chemical Abstract Service (CAS) numbers, analyte names, species names, and other coding conventions, as well as the lack of detection limits and associated data qualifiers (remark codes). The evaluation of toxicity data required the presence of control data. Depending on the data source, control data were not regularly reported with the data and could not be evaluated.

Some of the data analyzed for the tier classification were compiled as early as 1990 (the analyzed data cover the period of 1990-1999) and might not reflect current conditions. Emissions of many prominent contaminants have declined, and significant remediation efforts have taken place at many locations since that time. In addition, dredging, burial, natural attenuation, and scouring might have removed contaminants from some sampling stations. Unlike the first report to Congress, this analysis did include a temporal assessment of trends in sediment contaminant levels using data from 1980 through 1999, but it cannot be considered comprehensive and is applicable to only the locations where data were collected and evaluated.

Some data parameters are consistently absent throughout the NSI database. (Refer to Appendix A, Table A-1, for information on the number of NSI database stations at which the various types of data were compiled.) For example, only 10 percent of the stations with sediment chemistry data had associated toxicity data. For many of the fish tissue data included in the NSI database, the species was not identified. Also, additional assessment parameters (other than sediment chemistry, sediment toxicity, and tissue residue), such as benthic macroinvertebrate data, are not included in the NSI database and therefore not used in the evaluation process.

The unavailability of matching sediment chemistry and tissue residue data also limited the NSI data evaluation. In several instances, fish tissue was not analyzed for the same suite of chemicals for which sediment was analyzed. Spatial and temporal limitations of the data might have directly affected the analysis. Although some sediment chemistry and tissue residue data might have been collected in the same or very similar sampling stations, if the station names were not identical, the data could not be treated as if they were collected from the same location. This very likely resulted in an underestimate of the number of Tier 1 stations identified based on potential human health effects. The underestimate

occurred because exceedances of sediment TBP and tissue levels (EPA risk levels and FDA levels) at the same sampling station were required to categorize stations as Tier 1.

The lack of consistency among the different monitoring programs in the suite of chemicals analyzed also represents an area of uncertainty in the NSI data evaluation. Certain databases contain primarily information describing concentrations of metals or pesticides, whereas others contain data describing concentrations of nearly every chemical monitored in all of the NSI data. Many monitoring programs use a screening list of chemicals that are indicator pollutants for contaminated sediments. Thus, many of the specific chemicals assessed in the NSI data evaluation are not measured in every sample. In addition, certain classes of in-place sediment contaminants might not be recognized as causing significant impacts and thus are not routinely measured.

Information describing local background levels of sediment contaminants was usually not presented with the data included in the NSI database and thus was not considered when the significance of elevated contaminant concentrations in sediment was evaluated. Background conditions can be important in an evaluation of potential adverse effects on aquatic life because ecosystems can adapt to their ambient environmental conditions. For example, high metals concentrations in samples collected from a particular station might occur from natural geological conditions at that location, as opposed to the effects of human activities.

Most data are associated with a specific location and collected from a nonrandom sampling design. As a result, establishing the extent of contaminated sediment within a waterbody is not possible because it is difficult to assess the extent to which a monitoring station represents a larger segment of a waterbody. Furthermore, the NSI data are geographically biased. More than two-thirds of all stations evaluated in the NSI database are located in Washington, Virginia, California, Illinois, Florida, Wisconsin, New York, Texas, Oregon, and South Carolina. Each of these states has more than 500 monitoring stations. Finally, EPA did not verify reported latitude and longitude coordinates for each sampling station.

Limitations of Approach

Sediment Chemistry Screening Values

As was indicated in the first *National Sediment Quality Survey*, there are gaps in our knowledge concerning sediment-pollutant chemistry (especially bioavailability) and direct and indirect effects on aquatic biota. The certainty with which sediment toxicity can be predicted for each chemical using the various screening values included in the NSI database evaluation can vary significantly based on the quality of the available data and the appropriateness of exposure assumptions. For example, the draft ESGs are based on either secondary or final acute/chronic values which are not equivalent, even though they were developed using the same methodology. Draft ESGs based on final acute/chronic values are based on the highest quality toxicity and octanol/water partitioning data, which have been reviewed extensively. Draft ESGs based on secondary acute/chronic values have also undergone extensive field validation experiments. However, draft ESGs based on secondary acute/chronic values are in many cases based on a less extensive toxicity data set and have not been field validated.

The bioavailability of metals in sediment is addressed by the comparison of the molar concentration of sulfide anions (i.e., acid-volatile sulfide [AVS]) to the molar concentration of metals (i.e., simultaneously extracted metals [SEM]). To apply the [SEM]-[AVS] difference to indicate positive bioavailability and toxicity for this evaluation, EPA used laboratory data that indicated the probability of observed toxic effects at various [SEM]-[AVS] levels. Based on these data, EPA defined the Tier1 level as [SEM]-[AVS]>5. Thus, this use of [SEM]-[AVS] represents a hybrid of a theoretical approach and a correlative approach. Currently, the [SEM]-[AVS] difference is most usually considered an indicator of when metals are not bioavailable; however, some data have shown that metal bioaccumulation occurs where the [SEM]-[AVS] predicts no adverse effect. Differences in dietary exposures, applicability of equilibrium partitioning theory to sediment assessments, and varying redox conditions in some anaerobic

sediment might limit general applicability of the [SEM]-[AVS] method. Despite this limitations, EPA's Science Advisory Board (SAB, 2000) indicated that the [SEM]-[AVS] method "may be particularly useful to prioritize sites requiring attention ...."

Only those chemicals for which sediment chemistry screening values (i.e., draft ESGs) are available were evaluated in the analysis of NSI data. Therefore, the methodology could not identify contamination associated with chemical classes such as ionic organic compounds (e.g., alkyl phenols) and organometallic complexes (e.g., tributyl tin).

Biological effects correlation approaches such as the logistic model are based on the evaluation of paired field and laboratory data that relate adverse biological effects to the dry-weight chemical concentrations for a particular sample. Although computed from individual chemical observations, the predicted proportion toxic, it does not demonstrate that a particular chemical is solely responsible. In fact, a given sample typically contains a mixture of chemicals that contribute to observed adverse effects to some degree. For this reason, these correlative approaches are better at predicting toxicity in complex mixtures of contaminants in sediment.

Another concern is the application of screening values based on freshwater data (draft ESGs) and those based on saltwater data alone (logistic model) to evaluate sediment contaminant concentrations in the NSI database from both freshwater and saltwater habitats. Freshwater organisms exhibit tolerance to toxic chemicals similar to that of saltwater species when tested in their respective water; however, estuarine organisms might be less tolerant if osmotically stressed (Rand, 1995). Thus, the relative toxicity of a chemical in water (i.e., its chronic threshold water concentration) is usually within an order of magnitude for saltwater and freshwater species, although final chronic values and proposed sediment quality guidelines values are usually slightly higher for saltwater species. Ingersoll et al. (1996) reported similar reliability and predictive ability between marine and freshwater guidelines. The logistic model, as used in this assessment, was developed using only saltwater acute toxicity data.

Additional false positive and false negative classifications of risk to aquatic life from sediment contaminant concentrations could occur when a default value for organic carbon content is applied. Draft ESGs are based on the partitioning of a chemical between organic carbon in the sediment and pore water at equilibrium. Because the organic carbon content of most sediment samples in the NSI database is unknown, these sediment samples were assumed to contain 1 percent organic carbon. Total organic carbon (TOC) can range from 0.1 percent in sandy sediments to 1 to 4 percent in silty harbor sediments and 10 to 20 percent in navigation channel sediments (Clarke and McFarland, 1991). Long et al. (1995) reported an overall mean TOC concentration of 1.2 percent from data compiled from 350 publications for their biological effects database for sediments. Ingersoll et al. (1996) reported a mean TOC concentration of 2.7 percent with a 95 percent confidence interval of only 0.65 percent. In contrast, the concentration ranges of contaminants normalized to dry-weight typically varied by several orders of magnitude. Therefore, normalizing dry-weight concentrations to a relatively narrow range of TOC concentrations had little influence on relative concentrations of contaminants among samples.

Uncertainty associated with the equilibrium partitioning theory for developing draft ESGs includes the degree to which the equilibrium partitioning model explains the available sediment toxicity data (USEPA, 1993b). An analysis of variance using freshwater and saltwater organisms in water-only and sediment toxicity tests (using different sediments) was conducted to support development of the draft sediment guidelines. This analysis indicated that varying the exposure medium (i.e., water or sediment) resulted in an estimate of variability that should be used for computing confidence limits for the draft ESGs. The methodology used to derive the octanol/water partitioning coefficient and the final chronic value can also influence the degree of uncertainty associated with the draft ESGs. Differences in the response of water column and benthic organisms, as well as limitations in understanding the relationship of individual and population effects to community-level effects, have also been noted (Mancini and Plummer, 1994). Site-specific modifications to screening values derived using the equilibrium partitioning model have been recommended to better address chemical bioavailability and species

sensitivities (USEPA, 1993a). Sediment chemistry screening values developed using the equilibrium partitioning approach also do not address possible synergistic, antagonistic, or additive (except in the case of PAHs and metals as outlined in this chapter and Appendix B) effects of contaminants.

Fish Tissue Screening Values

The approach used to assess sediment chemistry data for the potential to accumulate in fish tissue also represents a theoretical approach with field-measured components. In addition to applying a site-specific or default organic carbon content, the TBP calculation includes a field-measured biota sediment accumulation factor (BSAF) to account for the relative affinity of a chemical for fish tissue lipids or sediment organic carbon. The BSAF will account for the effects of metabolism and biomagnification in the organism in which it is measured. The primary limitation of this approach is the applicability of a field-measured BSAF, or a percentile from a distribution of values, at a variety of sites where the conditions may vary.

TBPs were assumed to be equivalent to levels detectable in fish tissue. However, this approach might not completely account for biomagnification in the food chain, especially when using a BSAF derived from a benthic organism. In addition, it is assumed that sediment does not move, that contaminant sources other than sediment are negligible, that fish migration does not occur, and that exposure is consistent. The TBP calculation assumes that various lipids in different organisms and organic carbon in different sediments are similar and have distributional properties similar to the field-measured values used to derive BSAFs. Other simplifying assumptions are that chemicals are similarly exchanged between the sediments and tissues and that compounds behave alike, independent of site conditions other than organic carbon content. In reality, physical-chemical processes (e.g., diffusion through porous media and sediment mixing) can vary and limit the rate at which chemicals can exchange with bottom sediments. Uptake of contaminants by aquatic organisms is also a kinetic (rate-controlled) process that can vary and be slowed, for example, by awkward passage of a bulky molecule across biological membranes. Also, a BSAF of 1 (thermodynamic equilibrium) was used to estimate TBPs for many nonpolar organics. This BSAF might overestimate or underestimate the bioaccumulative potential for certain nonpolar organic chemicals because it is assumed that there is no metabolic degradation or biotransformation of such chemicals. Site-specific organic carbon content was often not available, which leads to additional uncertainty concerning the comparability of BSAFs among different locations. In addition, development of the BSAFs used in the TBP evaluation relied on a large amount of data that have not been published or peer-reviewed. Because of these factors, actual residue levels in fish resulting from direct and/or indirect exposure to contaminated sediment might be higher or lower. There is therefore uncertainty regarding sampling stations classifications based on comparison of estimated TBPs with FDA tolerance/action and guideline levels and EPA risk levels.

TBPs could not be calculated for polar organic compounds or heavy metals. Therefore, sampling stations could not be classified using FDA levels or EPA risk levels for those chemicals using a TBP approach (although fish tissue monitoring data are often available for many stations).

Uncertainties and numerous assumptions are associated with exposure parameters and toxicity data used to derive EPA risk levels and FDA tolerance/action and guideline levels. For example, the derivation of EPA risk levels is based on the assumption that an individual consumes on average 6.5 g/day of fish caught from the same site over a 70-year period. Also, the TBP calculation for human health assessments assumes fish tissue contains 3 percent lipid. This value is intended to be indicative of the fillet rather than the whole body. Generally, the exposure assumptions and safety factors incorporated into toxicity assessments might overestimate risks to the general population associated with sediment contamination but might underestimate risks to populations of subsistence or recreational fishers.

Whereas the Tier 1, Tier 2, and Tier 3 evaluation benchmarks established in this report represent recent advances in sediment assessment techniques, they have been utilized in this report as a way to relate all the different data from all the different sources around the U.S. using common benchmarks.

Therefore, the Tier 1, Tier 2, and Tier 3 benchmarks and interpretations used in this report are not currently appropriate for use in EPA regulatory programs that have developed their own frameworks and regulatory requirements, and were not designed to be a substitute for the various EPA program regulatory frameworks/authorities. EPA's regulatory programs (e.g., Office of Solid Waste and Emergency Response - OSWER) have developed their own scientifically defensible approaches to sediment evaluation based on the needs of their programs, and they will continue to utilize their current regulatory frameworks when making decisions regarding potentially contaminated sediments (e.g., sediment remediation, sediment disposal).

Other Limitations

Because a numerical score was not assigned to each sampling station to indicate the level of contamination associated with that station, it is not possible to determine which of the stations in Tier 1 should be considered the "most" contaminated. Such a numerical ranking system was intentionally not used for the NSI data evaluation because EPA does not believe that such ranking is appropriate for a screening-level analysis such as this, given the level of uncertainty.